A General Multi-Treatment Adaptive Design for Multivariate Responses

Atanu Biswas
Applied Statistics Unit, Indian Statistical Institute, Kolkata, India

D. Stephen Coad
School of Mathematical Sciences, Queen Mary, University of London, London, U.K.

Abstract: A general adaptive allocation design is proposed for continuous multivariate responses where the covariance matrix of the response vectors is unknown. There are $K > 2$ competing treatments, possible prognostic factors are considered in the allocation procedure, potential delayed responses are allowed for, and treatment-covariate interactions are incorporated. The allocation rule for any incoming patient is dependent on all the allocation-and-response-and-prognostic factor history of the previously allocated patients as well as the prognostic factor vector of the current patient. The design is a generalization of the approach suggested by Bandyopadhyay and Biswas (2001), which was presented for a much simpler scenario. The performance characteristics of the proposed design and some follow-up inference procedures are studied analytically and also numerically illustrated. An extension of the present approach to the situation where some components of the response vectors are continuous and some binary is then considered. Some further extensions of the work are briefly indicated.

Keywords: Allocation proportion; Delayed response; Mixed responses; Multivariate normal distribution; Probit link; Prognostic factor; Unknown variability.

Subject Classification: 62L05; 62H12; 62H15.

1. INTRODUCTION

The comparison of $K > 2$ treatments in a clinical trial has recently received considerable interest in the literature. Furthermore, several adaptive allocation designs have been proposed for assigning patients to the competing treatments. The aim of these designs is to allocate larger numbers of patients to the better treatments as the trial proceeds and thus to satisfy the ethical requirement that as few patients as possible are assigned to inferior ones. Of course, the simultaneous objective of estimating treatment differences is also under consideration. The idea of an adaptive design is to let the past data determine the allocation probabilities to the different treatments for the entering patients. However, most of the available literature on adaptive designs deals with treatments with binary responses, ignores associated prognostic factors for the incoming patients, and overlooks possible delays in responses. For example, Andersen et al. (1994) study an adaptive design for comparing $K > 2$ treatments with binary responses that are immediately ascertainable, and Ivanova and Rosenberger (2000) compare this design with several others.

One drawback of adaptive designs is that they are limited in terms of possible applications, due in part to the underlying models being unrealistic. For example, covariates may influence the responses, but almost all of the existing work on adaptive designs ignores such information. In the real-life applications described by Royall (1991) and Tamura et al. (1994), the available covariate information could not be used due to the unavailability of such theory. Depending on the covariate information, a patient's response may be more likely to be a success. A treatment-allocation problem in the presence of prognostic factors is considered by Begg and Iglewicz (1980), who used optimum-design theory to suggest a deterministic design criterion, which is then modified for computational convenience. The presence of prognostic factors is also considered by Atkinson (1982, 1999), who used optimum-design theory to provide a procedure based on the biased-coin design for an arbitrary number of treatments. Of course, the goal of biased-coin designs is to achieve equal allocation. In the context of adaptive designs, where the aim is to allocate larger numbers of patients to the better treatments, the treatment-allocation problem is discussed by Bandyopadhyay and Biswas (1999), who provided a modification of the randomized play-the-winner rule of Wei and Durham (1978) to incorporate covariate information. In the case of continuous responses,

such covariate information is used by Bandyopadhyay and Biswas (2001), and Rosenberger et al. (2001) consider the binary case.

As a motivating example for the type of design that we propose, consider the adaptive clinical trial that was conducted at the Indian Statistical Institute from January 1999 to March 2000, which is described by Biswas and Dewanji (2004). In this particular trial, the effect of pulsed electromagnetic-field therapy was compared with a placebo in the treatment of patients with rheumatoid arthritis. For each patient, there was a multivariate response vector, which included values on variables such as degree of pain and extent of swelling, in addition to various prognostic factors. Since no suitable adaptive design was available for multivariate responses, much of the information could not be used for treatment allocation. Moreover, the patients were assumed to be homogeneous, so that information on prognostic factors was also not incorporated.

In many practical situations, the responses are likely to be continuous. Incorporating the full continuous response history into the allocation design is a challenge to statisticians. Bandyopadhyay and Biswas (2001); henceforth referred to as BB, introduced a design in this direction, which may be viewed as a first attempt to skew the allocation probabilities when responses are continuous. They considered a simple linear model for univariate responses without having any treatment-covariate interaction in their model. However, their model is simple in the sense that, in reality, many more logistics come into play. A more realistic and general design in the following five respects is considered in the present paper.

1. BB considered the case of univariate responses from the patients. But, in reality, the responses are likely to be multivariate in many situations. For a recent account of some of the issues raised by multiple outcomes in the context of clinical trials, see chapter 15 of Jennison and Turnbull (2000). Also, in some applications, safety and efficacy may be the two components of a bivariate response; see, for example, Jennison and Turnbull (1993). In such cases, it may be a useful but difficult idea to use the multivariate response history for any adaptation. The present paper deals with this.
2. The approach of BB is for two treatments only. It is not straightforward to extend their method to a $K > 2$ treatment setup, as in the other multi-treatment generalizations of two-treatment adaptive designs. The present paper makes progress in this direction too, taking the multivariate responses into consideration.
3. One serious simplification in the approach of BB is the assumption of known variances for the errors in the linear model. The variances, a matrix in the present multivariate case, are likely to be unknown and

unequal for different components of the treatment response vector. This general and practical situation is considered in the present development. Consequently, the model and analysis change quite considerably.

4. BB considered a simple linear model and ignored any possible treatment-covariate interaction. The latter possibility is incorporated in the present study. The analysis of such interactions in the context of nonadaptive designs is discussed, for example, in chapter 7 of Whitehead (1997). In BB, only the allocation-and-response-and-prognostic factor history was considered in the adaptation. Thus, the current patient's prognostic factor vector was ignored. In the present work, we also use the current patient's prognostic factor for the adaptation, and the design is modified accordingly.
5. Finally, BB briefly mentioned possible delayed responses in the adaptation procedure. Here, we discuss this possibility in a more general way and present some discussions in this context. The effect of delayed responses in the context of binary data has been investigated by simulation by Ivanova and Rosenberger (2000). More recently, asymptotic properties of various adaptive urn designs for discrete data in the presence of delays are obtained by Bai et al. (2002).

The paper is structured as follows. In section 2, the general multivariate model is described, and it is shown how the parameters may be estimated. The adaptive design is introduced in section 3, and then some properties of the design are investigated in section 4 using simulation. This is followed in section 5 by a discussion of how inference may be carried out after such a design. The adaptive design in section 3 for continuous responses is then extended in section 6 to the case where some of the responses are binary and some continuous. In section 7, the main conclusions are summarized and some possible extensions to the present work are indicated.

2. THE GENERAL MODEL

2.1. Description of Model

In the present paper, we consider a general situation where we have multivariate treatment responses as well as $K > 2$ treatments. For the ith entering patient in the study, let $\underline{Y}_i$ be the $m \times 1$ multivariate response vector under consideration. Let us define a set $\{\delta_{1i}, \ldots, \delta_{Ki}\}$ of indicator variables such that $\delta_{ji} = 1$ if the ith patient is treated by treatment j and $\delta_{ji} = 0$ otherwise. Clearly, $\sum_{j=1}^{K} \delta_{ji} = 1$ for all i. Let $\underline{\mu}_j$ be the $m \times 1$ treatment effect for the jth treatment for $j = 1, \ldots, K$. Further, suppose that $\underline{x}_i$ is the $p \times 1$ vector of prognostic factors for the ith

entering patient and that $B_j = (\underline{\beta}_{j1}, \ldots, \underline{\beta}_{jm})^T$ is the corresponding $m \times p$ matrix of regression coefficients. Then we can write the linear model for multivariate response as

$$\underline{Y}_i = \sum_{j=1}^{K} \delta_{ji} \underline{\mu}_j + \sum_{j=1}^{K} \delta_{ji} B_j \underline{x}_i + \underline{\epsilon}_i, \tag{2.1}$$

where $\underline{\epsilon}_i$ is the $m \times 1$ error vector. Note that the model includes both the main prognostic factors and their interactions with the treatments, whereas the multivariate analogue of the BB model would have $B_j = B$ for all j. In the present paper, we carry out the analysis assuming independent multivariate distributions for the $\underline{\epsilon}_i$s with mean vector zero and an unknown covariance matrix Σ. The design provided in section 3 is applicable for any distribution of $\underline{\epsilon}_i$s. Note that if the δ_{ji}s are predetermined and the errors are normal, (2.1) is just the usual multivariate analysis of covariance model.

Now, the general model (2.1) can be written in the form

$$Y_n = Z_n \theta + \epsilon_n, \tag{2.2}$$

where

$$Y_n = (\underline{Y}_1, \ldots, \underline{Y}_n)^T, \quad \theta = (\underline{\mu}_1, \ldots, \underline{\mu}_K, B_1, \ldots, B_K)^T, \quad \epsilon_n = (\underline{\epsilon}_1, \ldots, \underline{\epsilon}_n)^T$$

and

$$Z_n = \begin{pmatrix} \underline{\delta}_1^T & \delta_{11}\underline{x}_1^T & \ldots & \delta_{K1}\underline{x}_1^T \\ \underline{\delta}_2^T & \delta_{12}\underline{x}_2^T & \ldots & \delta_{K2}\underline{x}_2^T \\ \vdots & \vdots & & \vdots \\ \underline{\delta}_n^T & \delta_{1n}\underline{x}_n^T & \ldots & \delta_{Kn}\underline{x}_n^T \end{pmatrix}.$$

Note that $\underline{\delta}_i = (\delta_{1i}, \ldots, \delta_{Ki})^T$ for $i = 1, \ldots, n$.

2.2. Estimation of Parameters

From (2.2), the normal equations are

$$Z_n^T Z_n \theta = Z_n^T Y_n, \tag{2.3}$$

where

$$Z_n^T Z_n = \begin{pmatrix} \sum_{i=1}^n \underline{\delta}_i \underline{\delta}_i^T & \sum_{i=1}^n \delta_{1i} \underline{\delta}_i \underline{x}_i^T & \ldots & \sum_{i=1}^n \delta_{Ki} \underline{\delta}_i \underline{x}_i^T \\ \sum_{i=1}^n \delta_{1i} \underline{x}_i \underline{\delta}_i^T & \sum_{i=1}^n \delta_{1i}^2 \underline{x}_i \underline{x}_i^T & \ldots & \sum_{i=1}^n \delta_{1i}\delta_{Ki} \underline{x}_i \underline{x}_i^T \\ \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^n \delta_{Ki} \underline{x}_i \underline{\delta}_i^T & \sum_{i=1}^n \delta_{1i}\delta_{Ki} \underline{x}_i \underline{x}_i^T & \ldots & \sum_{i=1}^n \delta_{Ki}^2 \underline{x}_i \underline{x}_i^T \end{pmatrix}$$

and

$$Z_n^T Y_n = \begin{pmatrix} \sum_{i=1}^n \underline{\delta}_i \underline{Y}_i^T \\ \sum_{i=1}^n \delta_{1i} \underline{x}_i \underline{Y}_i^T \\ \vdots \\ \sum_{i=1}^n \delta_{Ki} \underline{x}_i \underline{Y}_i^T \end{pmatrix}.$$

Note that $\delta_{ji}\delta_{j'i} = 1$ or 0 according to whether $j = j'$. Now let

$$N_{jn} = \sum_{i=1}^n \delta_{ji}, \qquad \overline{\underline{Y}}_{jn} = \frac{1}{N_{jn}} \sum_{i=1}^n \delta_{ji} \underline{Y}_i$$

and

$$\overline{\underline{x}}_{jn} = \frac{1}{N_{jn}} \sum_{i=1}^n \delta_{ji} \underline{x}_i$$

for $j = 1, \ldots, K$. Then (2.3) implies that

$$N_{jn} \underline{\mu}_j + N_{jn} B_j \overline{\underline{x}}_{jn} = N_{jn} \overline{\underline{Y}}_{jn} \tag{2.4}$$

and

$$N_{jn} \overline{\underline{x}}_{jn} \underline{\mu}_j^T + \sum_{i=1}^n \delta_{ji} \underline{x}_i \underline{x}_i^T B_j^T = \sum_{i=1}^n \delta_{ji} \underline{x}_i \underline{Y}_i^T \tag{2.5}$$

for $j = 1, \ldots, K$. Solving (2.4) and (2.5) yields the estimated covariate-adjusted treatment effect for the jth treatment given by

$$\hat{\underline{\mu}}_{jn} = \overline{\underline{Y}}_{jn} - \widehat{B}_{jn} \overline{\underline{x}}_{jn}, \tag{2.6}$$

where

$$\widehat{B}_{jn}^T = \left(\sum_{i=1}^n \delta_{ji} \underline{x}_i \underline{x}_i^T - N_{jn} \overline{\underline{x}}_{jn} \overline{\underline{x}}_{jn}^T \right)^{-1} \left(\sum_{i=1}^n \delta_{ji} \underline{x}_i \underline{Y}_i^T - N_{jn} \overline{\underline{x}}_{jn} \overline{\underline{Y}}_{jn}^T \right)$$

for $j = 1, \ldots, K$. Now write

$$\widehat{B}_{jn}^T = (\hat{\underline{\beta}}_{j1n}, \ldots, \hat{\underline{\beta}}_{jmn})$$

for $j = 1, \ldots, K$. Then it follows from (2.6) that the estimated covariate-adjusted difference between treatments j and k in the ℓth response component is

$$\hat{\mu}_{j\ell n} - \hat{\mu}_{k\ell n} = \overline{Y}_{j\ell n} - \overline{Y}_{k\ell n} - \hat{\underline{\beta}}_{j\ell n}^T \overline{\underline{x}}_{jn} + \hat{\underline{\beta}}_{k\ell n}^T \overline{\underline{x}}_{kn}$$

for $j, k = 1, \ldots, K$, $j \neq k$ and $\ell = 1, \ldots, m$. Note that the forms of the above estimators are not affected by the adaptive design, but their distributions are.

2.3. Delayed Responses

Let η_{un} be an indicator variable which takes the value 1 or 0 according to whether the response of the uth patient is available just before the entry of the nth patient. Thus, $\{\eta_{un};\ u = 1, \ldots,\ n = u+1, \ldots\}$ is a triangular sequence such that any row is either $\{0, \ldots, 0, \ldots\}$, $\{1, \ldots, 1, \ldots\}$, or $\{0, \ldots, 0, 1, \ldots, 1, \ldots\}$. We write

$$P(\eta_{un} = 1) = \pi_{n-u},$$

where $\pi_{n+1} \geq \pi_n$ for all n and $\pi_n \to 1$ as $n \to \infty$. This means that every response is obtainable eventually. It is easy to see that, for $t > s$,

$$P(\eta_{ut} = 1 | \eta_{us} = 1) = 1,$$

but we also need to model $P(\eta_{ut} = 1 | \eta_{us} = 0)$. A useful functional form for this probability may be $1 - a^{t-s}$, for some $a \in (0, 1)$, which can be justified from a mathematical formulation considering some distribution of the response times, some other distribution for the interarrival times, and independence of these distributions; see, for example, Bandyopadhyay and Biswas (1996).

When some responses are delayed, the delayed response indicators η_{un}s come into play. In fact, if some responses are not available, the convenient way is to carry out the analysis using the available responses, provided the available responses are sufficient in number to find estimates of all of the unknown parameters. Thus, the approach will be conditional on the η_{un}s. For mathematical simplicity, we assume that the η_{un}s are independent of any other variables, that is, the Ys and δs, which is practical in many situations. In this case, by using a superscript D for delayed response estimates, the estimates of the components of θ will be

$$\underline{\hat{\mu}}_{j0}^{D} = \underline{\overline{Y}}_{j0}^{D} - \widehat{B}_{j0}^{D} \underline{\overline{x}}_{j0}^{D},$$

where

$$N_{j0}^{D} = \sum_{i=1}^{n} \eta_{i,n+1} \delta_{ji}, \quad \underline{\overline{Y}}_{j0}^{D} = \frac{1}{N_{j0}^{D}} \sum_{i=1}^{n} \eta_{i,n+1} \delta_{ji} \underline{Y}_i,$$

$$\underline{\overline{x}}_{j0}^{D} = \frac{1}{N_{j0}^{D}} \sum_{i=1}^{n} \eta_{i,n+1} \delta_{ji} \underline{x}_i,$$

and

$$(\widehat{B}_{j0}^{D})^T = \left\{ \sum_{i=1}^{n} \eta_{i,n+1} \delta_{ji} \underline{x}_i \underline{x}_i^T - N_{j0}^{D} \underline{\overline{x}}_{j0}^{D} (\underline{\overline{x}}_{j0}^{D})^T \right\}^{-1}$$

$$\times \left\{ \sum_{i=1}^{n} \eta_{i,n+1} \delta_{ji} \underline{x}_i \underline{Y}_i^T - N_{j0}^{D} \underline{\overline{x}}_{j0}^{D} (\underline{\overline{Y}}_{j0}^{D})^T \right\}$$

for $j = 1, \ldots, K$. Here, $\overline{Y}_{j\cdot n}^{D}$ and $\overline{x}_{j\cdot n}^{D}$ are defined if the N_{jn}^{D} are nonzero, and $(\hat{B}_{jn}^{D})^{T}$ is defined if $\left\{ \sum_{i=1}^{n} \eta_{i,n+1} \delta_{ji} \underline{x}_i \underline{x}_i^{T} - N_{jn}^{D} \overline{\underline{x}}_{j\cdot n}^{D} (\overline{\underline{x}}_{j\cdot n}^{D})^{T} \right\}$ is nonsingular. An unconditional analysis, integrating over the distribution of the delayed responses, is of interest, but it is very difficult to proceed theoretically, as the η_{un}s are dependent for a fixed u. However, we provide some simulation results in section 4.2 that are unconditional on the delayed responses.

3. THE ADAPTIVE DESIGN

Let us denote by $\mathcal{F}_n$ the allocation, response, and prognostic factor histories up to the nth patient. Then, by our adaptive design, the allocation probabilities to the K competing treatments should be a function of $\mathcal{F}_n$, in some way. In the univariate two-treatment setup of Bandyopadhyay and Biswas (2001), the estimates of the components of $\underline{\mu}_j$ were based on these past data and these estimates were used for any adaptation. Here, we will use the same principle. In addition, one would like to use the prognostic factor of the entering patient, which is known before the allocation, in the adaptive design. In the case of a single binary response variable, this approach was taken by Rosenberger et al. (2001). Furthermore, there is a significant difference in our model from the approach of Bandyopadhyay and Biswas (2001) in the sense that we assume that the variance is unknown, and thus we need to make an appropriate adjustment to take this into account. Let $\hat{\sigma}_{\ell n}^{2}$ be the estimate of the error variance for the ℓth variable based on the data in $\mathcal{F}_n$, where

$$\{n - K(p+1)\}\hat{\sigma}_{\ell n}^{2} = \sum_{i=1}^{n} \sum_{j=1}^{K} \delta_{ji} \left(Y_{j\ell i} - \hat{\mu}_{j\ell n} - \underline{\hat{\beta}}_{j\ell n}^{T} \underline{x}_i \right)^{2}$$

for $\ell = 1, \ldots, m$.

We assign the first Km_0 patients so that exactly m_0 patients receive each treatment. Choice of $m_0 = m_0(N)$, where N is the total sample size, is such that m_0 is increasing with N, and an initial estimate of all the unknown parameters is obtainable from this initial sample of size Km_0. This initial sample can be looked upon as a compromise between adaptive and fixed sample-size procedures. It ensures that at least m_0 patients are assigned to each treatment and that we have some estimates of the unknown parameters as, subsequently, some treatments may receive very few, or even no, allocations using the adaptive procedure. Of course, if m_0 is too large, there will be little skewing of the allocation probabilities, since the adaptive procedure is only applied to the last $(N - Km_0)$ patients. From the $(Km_0 + 1)$st patient onwards, we carry out adaptive sampling in the following way.

We consider a suitable distribution function G of a symmetric random variable such that $G(0) = 1/2$ and $G(-x) = 1 - G(x)$. Then we set the allocation probability for the $(n+1)$st patient as

$$P(\delta_{j,n+1} = 1 | \mathcal{F}_n, \underline{x}_{n+1}) = \frac{1}{\binom{K}{2}} \sum_{k=1(k\neq j)}^{K} \sum_{\ell=1}^{m} w_\ell \times G\left(\frac{\hat{\mu}_{j\ell n} + \underline{x}_{n+1}^T \hat{\underline{\beta}}_{j\ell n} - \hat{\mu}_{k\ell n} - \underline{x}_{n+1}^T \hat{\underline{\beta}}_{k\ell n}}{\hat{\sigma}_{\ell n}} \right), \quad (3.1)$$

where w_ℓ is the weight for the ℓth component of the multivariate response with $\sum_{\ell=1}^{m} w_\ell = 1$. Note that (3.1) is an example of what Rosenberger et al. (2001) calls a treatment effect mapping. It was first used by Rosenberger (1993) for continuous responses. The choice of w_ℓ is, of course, a delicate question. Unlike in the usual sampling case, w_ℓ should not be variance-driven, since here the ethical issue is of greater concern and the more important response component should receive greater weight in the allocation procedure. In section 4.1, we study the effect of different choices of the w_ℓ upon the allocation proportions using simulation. In the case of possible delayed responses, (3.1) will be replaced by

$$P(\delta_{j,n+1} = 1 | \mathcal{F}_n, \underline{x}_{n+1}) = \frac{1}{\binom{K}{2}} \sum_{k=1(k\neq j)}^{K} \sum_{\ell=1}^{m} w_\ell \times G\left(\frac{\hat{\mu}^D_{j\ell n} + \underline{x}_{n+1}^T \hat{\underline{\beta}}^D_{j\ell n} - \hat{\mu}^D_{k\ell n} - \underline{x}_{n+1}^T \hat{\underline{\beta}}^D_{k\ell n}}{\hat{\sigma}^D_{\ell n}} \right).$$

4. NUMERICAL RESULTS

4.1. Immediate Responses

We carried out a detailed simulation study with $K = 3$ treatments for bivariate responses where both components of the response are equally important, that is, $w_1 = w_2 = 0.5$, and also for the cases where $w_1 = 0.8$ and 0.2. Further, we consider the case $p = 1$, for illustration. We took $m_0 = 4$, so that the adaptive design is applied to the 13th patient onwards: see section 3. In the computations, we used the probit link with the history $G(x) = \Phi(x)$, where Φ denotes the standard normal distribution function; we have taken $\sigma_1^2 = \sigma_2^2 = 4$ and $B_1 = B_2 = B_3 = (1, 2)^T$, and ρ denotes the correlation between the responses and $X \sim N(2, 1)$. The results for $n = 30$ and $n = 60$ are based on 1,000

simulations. In both cases, the results were obtained using a computer program written in S-Plus. The allocation proportions and their standard deviations when $w_1 = w_2 = 0.5$ are reported in Table 1. We observe that a larger number of patients are treated by the better treatment by our allocation design, as expected.

To see the effect of the two components of the response not being equally important, a simulation study as above was carried out when $\underline{\mu}_1 = (3, 2)^T$, $\underline{\mu}_2 = (2, 2)^T$, $\underline{\mu}_3 = (1, 2)^T$, $n = 60$, and $\rho = 0.5$. When $w_1 = 0.5$, the allocation proportions were 0.407, 0.325, and 0.268 with standard deviations 0.131, 0.137, and 0.132. When $w_1 = 0.8$, so that the first component is more important, these figures become 0.433, 0.336, and

Table 1. Monte Carlo estimates of means and standard deviations of allocation proportions

n	ρ	$E(N_{1n}/n)$	$SD(N_{1n}/n)$	$E(N_{2n}/n)$	$SD(N_{2n}/n)$	$E(N_{3n}/n)$	$SD(N_{3n}/n)$
(a) $\underline{\mu}_1 = \underline{\mu}_2 = \underline{\mu}_3 = (2, 2)^T$							
30	0.5	0.333	0.120	0.333	0.120	0.334	0.120
30	0.1	0.333	0.110	0.333	0.110	0.334	0.110
60	0.5	0.334	0.140	0.333	0.140	0.333	0.140
60	0.1	0.333	0.130	0.334	0.130	0.333	0.130
(b) $\underline{\mu}_1 = (3, 3)^T$ and $\underline{\mu}_2 = \underline{\mu}_3 = (2, 2)^T$							
30	0.5	0.389	0.120	0.301	0.115	0.310	0.114
30	0.1	0.388	0.105	0.307	0.103	0.305	0.103
60	0.5	0.416	0.120	0.292	0.140	0.292	0.140
60	0.1	0.433	0.122	0.285	0.128	0.281	0.121
(c) $\underline{\mu}_1 = (3, 3)^T$, $\underline{\mu}_2 = (2, 2)^T$ and $\underline{\mu}_3 = (1, 1)^T$							
30	0.5	0.409	0.119	0.331	0.123	0.260	0.106
30	0.1	0.409	0.108	0.334	0.110	0.258	0.094
60	0.5	0.463	0.124	0.331	0.136	0.206	0.116
60	0.1	0.455	0.116	0.333	0.122	0.212	0.104
(d) $\underline{\mu}_1 = (4, 4)^T$, $\underline{\mu}_2 = (2, 2)^T$ and $\underline{\mu}_3 = (1, 1)^T$							
30	0.5	0.463	0.104	0.301	0.105	0.236	0.094
30	0.1	0.454	0.095	0.303	0.100	0.243	0.088
60	0.5	0.519	0.109	0.289	0.126	0.192	0.105
60	0.1	0.520	0.099	0.286	0.118	0.194	0.094
(e) $\underline{\mu}_1 = (4, 4)^T$, $\underline{\mu}_2 = (2, 2)^T$ and $\underline{\mu}_3 = (0, 0)^T$							
30	0.5	0.468	0.100	0.327	0.118	0.205	0.078
30	0.1	0.464	0.090	0.333	0.095	0.203	0.073
60	0.5	0.533	0.097	0.330	0.100	0.137	0.083
60	0.1	0.536	0.078	0.324	0.097	0.140	0.071

0.231 with standard deviations 0.141, 0.142, and 0.125. In contrast, when $w_1 = 0.2$, so that the second component is more important, the allocation proportions are 0.363, 0.335, and 0.302 with standard deviations 0.149, 0.147, and 0.149. Clearly, these results make sense given the values for $\underline{\mu}_1$, $\underline{\mu}_2$ and $\underline{\mu}_3$.

4.2. Delayed Responses

We have carried out a detailed simulation study for delayed responses too and the allocation proportions are reported in Table 2. In the computations, we assume that sufficient time has passed in order to observe all of the responses from the first 12 patients, so that the delayed responses are applicable from the 13th patient onwards. Here, we used the following model. The probability of obtaining a response from the nth patient before the entry of the $(n+1)$st patient is $1 - \exp(-0.3)$. Further, the probability of obtaining a response from the nth patient before the entry of the $(n+t+1)$st patient, given that the response is not available before the entry of the $(n+t)$th patient, is $1 - \exp(-0.3)$. To appreciate the effect of this delayed response mechanism, if $n = 30$,

Table 2. Monte Carlo estimates of means and standard deviations of allocation proportions when there are delayed responses

n	ρ	$E(N_{1n}/n)$	$SD(N_{1n}/n)$	$E(N_{2n}/n)$	$SD(N_{2n}/n)$	$E(N_{3n}/n)$	$SD(N_{3n}/n)$
(a) $\underline{\mu}_1 = \underline{\mu}_2 = \underline{\mu}_3 = (2, 2)^T$							
30	0.5	0.333	0.110	0.333	0.110	0.334	0.110
60	0.5	0.333	0.140	0.333	0.140	0.334	0.140
(b) $\underline{\mu}_1 = (3, 3)^T$ and $\underline{\mu}_2 = \underline{\mu}_3 = (2, 2)^T$							
30	0.5	0.377	0.110	0.312	0.112	0.311	0.109
60	0.5	0.402	0.141	0.304	0.136	0.294	0.132
(c) $\underline{\mu}_1 = (3, 3)^T$, $\underline{\mu}_2 = (2, 2)^T$ and $\underline{\mu}_3 = (1, 1)^T$							
30	0.5	0.399	0.112	0.336	0.110	0.266	0.108
60	0.5	0.447	0.138	0.336	0.127	0.216	0.118
(d) $\underline{\mu}_1 = (4, 4)^T$, $\underline{\mu}_2 = (2, 2)^T$ and $\underline{\mu}_3 = (1, 1)^T$							
30	0.5	0.438	0.107	0.312	0.107	0.250	0.092
60	0.5	0.509	0.113	0.293	0.113	0.198	0.118
(e) $\underline{\mu}_1 = (4, 4)^T$, $\underline{\mu}_2 = (2, 2)^T$ and $\underline{\mu}_3 = (0, 0)^T$							
30	0.5	0.449	0.099	0.329	0.101	0.222	0.082
60	0.5	0.528	0.096	0.322	0.109	0.150	0.081

then about 26 responses on average are observed before the entry of the 31st patient.

As expected, we observe that the allocation proportions for such a delayed response case are weighted averages of the proportions obtained for immediate responses and $1/K$, and thus are slightly more skewed towards $1/K$ than the immediate response case. This can be easily explained, as the responses are possibly delayed and the adaptation will be slower, resulting in a lower skewness of the allocation probabilities.

5. INFERENCE FOLLOWING THE DESIGN

So far, we have mainly concentrated on design issues. Of course, after the adaptive allocation is complete, the question of relevant inference comes into play. To appreciate some of the difficulties that can arise, see Ivanova et al. (2000), who study an adaptive urn design for comparing $K > 2$ treatments with binary responses. Given the generality of our model, inference in our case is a separate and detailed issue. However, in this section, we provide a sketch of methods of inference. More specifically, we present some simulation results in order to assess how the conventional methods are affected by the adaptive design.

Suppose that, upon completion of the trial, we have data on n patients. Then there are a number of hypotheses that we may be interested in. For our purposes, the one of most interest is likely to be whether the treatment mean vectors are equal. The total sum of squares and products matrix for the responses is given by

$$T_{yy}^{(n)} = \sum_{j=1}^{K} \sum_{i=1}^{n} \delta_{ji} \underline{Y}_i \underline{Y}_i^T - n \overline{\underline{Y}}_n \overline{\underline{Y}}_n^T,$$

where

$$\overline{\underline{Y}}_n = \frac{1}{n} \sum_{j=1}^{K} N_{jn} \overline{\underline{Y}}_{jn}.$$

We may define $T_{xx}^{(n)}$ and $T_{xy}^{(n)}$ similarly. Continuing, the within-treatments sum of squares and products matrix for the responses is

$$W_{yy}^{(n)} = \sum_{j=1}^{K} \sum_{i=1}^{n} \delta_{ji} \underline{Y}_i \underline{Y}_i^T - \sum_{j=1}^{K} N_{jn} \overline{\underline{Y}}_{jn} \overline{\underline{Y}}_{jn}^T.$$

Again, we define $W_{xx}^{(n)}$ and $W_{xy}^{(n)}$ in the same way. Then the total sum of squares and products matrix for the responses, adjusted for the presence of the covariates, is

$$T_{yy.x}^{(n)} = T_{yy}^{(n)} - \{T_{xy}^{(n)}\}^T \{T_{xx}^{(n)}\}^{-1} T_{xy}^{(n)},$$

and $W_{yy.x}^{(n)}$ may be defined similarly. So the usual Wilks' lambda statistic for testing whether there is a difference in the treatment mean vectors is given by

$$\Lambda_n = \frac{|W_{yy.x}^{(n)}|}{|T_{yy.x}^{(n)}|}.$$

Note that the form of the statistic is not affected by the adaptive design, but its distribution is. However, although Λ_n no longer follows a Wilks' lambda distribution, the power of the exact test based on Λ_n can be evaluated using simulation. Of course, if we ignore the adaptive design, an approximate critical region can be constructed by using Bartlett's approximation for large n. Thus, in our case, if we ignore the adaptive design, we have

$$-\left\{n - \frac{1}{2}(m + K + 2p + 2)\right\} \log \Lambda_n \sim \chi^2_{m(K-1)}$$

asymptotically as $n \to \infty$, and it is interesting to compare the powers of these two tests.

Another hypothesis that we may be interested in is whether the treatment mean vectors are related by a linear contrast, such as $\sum_{j=1}^{K} \gamma_j \underline{\mu}_j = 0$, where $\gamma_1, \ldots, \gamma_K$ are fixed constants and $\sum_{j=1}^{K} \gamma_j = 0$. Clearly, the contrast sum of squares and products matrix for the responses is given by

$$C_{yy}^{(n)} = \frac{(\sum_{j=1}^{K} \gamma_j \overline{\underline{Y}}_{jn})(\sum_{j=1}^{K} \gamma_j \overline{\underline{Y}}_{jn}^T)}{\sum_{j=1}^{K} \frac{\gamma_j^2}{N_{jn}}}.$$

Thus, using chapter 12 of Mardia et al. (1979), the usual Wilks' lambda statistic for testing for the above linear contrast across the treatment groups is

$$\Lambda_n^c = \frac{|W_{yy.x}^{(n)}|}{|W_{yy.x}^{(n)} + C_{yy}^{(n)}|}.$$

As for the previous test, if we ignore the adaptive design, an approximate critical region can be constructed by using Bartlett's approximation for large n. In this case, if we ignore the adaptive design, we have

$$-\left\{n - \frac{1}{2}(m + 2K + 2p)\right\} \log \Lambda_n^c \sim \chi^2_m$$

asymptotically as $n \to \infty$.

Monte Carlo estimates of the powers of both the exact and approximate tests based on Λ_n at the 5% level of significance are presented in Table 3, where, in order to make the comparison of the two tests fair, the critical region of the approximate test is adjusted to give a test of size 5%. As before, the estimates are based on 1,000 simulations. Note that the exact test based on Λ_n is a left-tailed test while the above approximate test is right-tailed. These results show that the approximate test can have substantially less power than the corresponding exact test, especially for scenarios (b) and (c). Further simulations, not reported here, show that the size of the approximate test using an unadjusted critical region tends to be less than 5% for high values of ρ and greater than 5% for low values of ρ. In general, this test maintains the desired size of 5% if n is at least 50 in the present context.

If the hypothesis of equality of treatment mean vectors is rejected, then we will want to identify which differences are significantly different. Thus, we may be interested in the pairwise differences $\mu_{j\ell} - \mu_{k\ell}$ for $j, k = 1, \ldots, K$, $j < k$, and $\ell = 1, \ldots, m$. Covariate-adjusted estimators of these are given in section 2.2. If we again ignore the adaptive design, we can construct $100(1-\alpha)\%$ simultaneous confidence intervals for the pairwise differences using

$$\hat{\mu}_{j\ell n} - \hat{\mu}_{k\ell n} \pm t_{n-K(p+1),\alpha/\{mK(K-1)\}}\,\hat{\sigma}_{\ell n}\sqrt{\frac{1}{N_{jn}} + \frac{1}{N_{kn}} + \underline{\bar{x}}_{jn}^{T} S_{j\cdot x\cdot x}^{(n)} \underline{\bar{x}}_{jn} + \underline{\bar{x}}_{kn}^{T} S_{knn}^{(n)} \underline{\bar{x}}_{kn}}$$

for $j, k = 1, \ldots, K$, $j < k$, and $\ell = 1, \ldots, m$, where $t_{v,\gamma}$ represents the upper $\gamma\%$ point of the t distribution on v degrees of freedom and

$$S_{j\cdot x\cdot x}^{(n)} = \sum_{i=1}^{n} \delta_{ji} (\underline{x}_i - \underline{\bar{x}}_{jn})(\underline{x}_i - \underline{\bar{x}}_{jn})^T.$$

Table 3. Monte Carlo estimates of the powers at the 5% level for the exact and approximate tests based on Λ_n for $n = 30$ and 60

			Scenario				
n	ρ	Test	(a)	(b)	(c)	(d)	(e)
30	0.5	Exact	0.05	0.176	0.419	0.776	0.901
		Approx.	0.05	0.171	0.393	0.753	0.892
30	0.1	Exact	0.05	0.201	0.481	0.910	0.981
		Approx.	0.05	0.136	0.377	0.856	0.965
60	0.5	Exact	0.05	0.289	0.663	0.962	0.986
		Approx.	0.05	0.194	0.525	0.940	0.980
60	0.1	Exact	0.05	0.479	0.844	0.996	0.998
		Approx.	0.05	0.355	0.784	0.996	0.998

Table 4. Overall coverage probabilities for the simultaneous confidence intervals when $\alpha = 0.05$

		Scenario				
n	ρ	(a)	(b)	(c)	(d)	(e)
30	0.5	0.864	0.850	0.852	0.866	0.870
30	0.1	0.873	0.865	0.865	0.866	0.883
60	0.5	0.900	0.890	0.906	0.918	0.920
60	0.1	0.930	0.922	0.934	0.930	0.946

Simulation estimates of the overall coverage probabilities of these intervals based on 1,000 replications are given in Table 4. As expected, the coverage probabilities become closer to the nominal value of 95% as n increases and ρ decreases. However, even for $n = 60$ and $\rho = 0.1$, these can be up to 3% too small.

6. MIXED CONTINUOUS AND BINARY RESPONSES

In this section, for the purposes of illustration, we consider the bivariate case only, where one component of the response is continuous, assumed to be normally distributed, and the other component is binary. Such mixed continuous and discrete data, that is, binary/categorical responses, are common in several biomedical contexts. For some examples of such mixed biomedical data in the recent literature, see Holmes et al. (1994) for the Boston Convulsant Teratogenesis Study at Brigham and Women's Hospital, Wang et al. (1994) for the Harvard Six Cities Study, and Little and Schluchter (1985) for an example on the effects of parental psychological disorders on child development. For such a bivariate response, we can use a combination of our present approach, given in (3.1), and a K-treatment generalization of that of Bandyopadhyay and Biswas (1999) for binary responses in the two-treatment case with prognostic factors. The conditional allocation probability of the $(n+1)$st patient to the jth treatment will have a weight w_1 for the continuous response part and a weight $w_2 = 1 - w_1$ for the binary response part. The continuous part will be as given in section 3.

For the binary part, let g be a function that transforms $\mathcal{R}^p \to [0, H] \in \mathcal{R}$ such that $g(\underline{x}_1) > g(\underline{x}_2)$ when $\underline{x}_1$ is a more favorable prognostic factor vector than $\underline{x}_2$ in terms of treating a patient. In the expression for the conditional probability, the contribution of the binary responses will be a mixture of several parts. Let α_1 and α_2 be positive constants, and let $b_n = K + (K-1)\alpha_1 n + H\alpha_2 n$. Then each of the earlier

successes by the same treatment, that is, treatment j, will have a weight $\alpha_1(K-1)/b_n$; each of the earlier failures by any other treatment will have a weight α_1/b_n; and each of the earlier patients having a prognostic factor $\underline{x}$ will have a weight $\alpha_2\{H-g(\underline{x})\}/b_n$ if they are treated by treatment j and will have a weight $g(\underline{x})/\{(K-1)b_n\}$ if they are treated by some treatment other than j. Thus, the conditional probability of allocation to the jth treatment for the $(n+1)$st patient is

$$
\begin{aligned}
&P(\delta_{j,n+1}=1|\mathcal{F}_n,\underline{x}_{n+1})\\
&\quad=\frac{w_1}{\binom{K}{2}}\sum_{k=1(k\neq j)}^{K}G\left(\frac{\hat{\mu}_{jn}+\underline{x}_{n+1}^T\hat{\underline{\beta}}_{jn}-\hat{\mu}_{kn}-\underline{x}_{n+1}^T\hat{\underline{\beta}}_{kn}}{\hat{\sigma}_{1n}}\right)\\
&\quad+\frac{w_2}{b_n}\left(1+\alpha_1\left\{(K-1)\sum_{i=1}^{n}Y_{2i}\delta_{j\bar{i}}+\sum_{i=1}^{n}(1-Y_{2i})(1-\delta_{j\bar{i}})\right\}\right.\\
&\quad\left.+\alpha_2\left[\sum_{i=1}^{n}\{H-g(\underline{x}_i)\}\delta_{ji}+\sum_{i=1}^{n}g(\underline{x}_i)(1-\delta_{ji})/(K-1)\right]\right).
\end{aligned}
$$

Note that, by this allocation design, the current patient's prognostic factor is only used in the continuous response part. In order to use the current patient's prognostic factor in the binary response part, a logistic response function could be used, as in Rosenberger et al. (2001).

The performance of the design is illustrated using a simple simulation study. We considered the case of $K=3$ treatments for bivariate responses, where one component is continuous and the other binary, and both components are equally important. As in section 4.1, $p=1$, $G(x)=\Phi(x)$, $\sigma_1^2=4$, $m_0=4$, and $X\sim N(2,1)$. Further, $\alpha_1=\alpha_2=1$, $B_1=B_2=B_3=1$, $H=1$, and $g(x)=1_{\{x>2\}}$, where 1_A denotes the indicator of the event A. Let p_1, p_2, and p_3 denote the success probabilities for the binary component for the three treatments when $g(x)=1$. When $g(x)=0$, these success probabilities are a_0p_1, a_0p_2, and a_0p_3. The results for $n=60$ based on 1,000 simulations are presented in Table 5 when $a_0=0.75$. As expected, we see that a larger number of patients are assigned to the better treatment by our allocation design.

7. DISCUSSION

A general adaptive design has been proposed in this paper for use when several treatments with continuous multivariate responses are to be compared. Simulation studies indicate that the design successfully assigns more patients to the better treatments. We have also shown how delayed responses may be incorporated, studied the power of the design

Table 5. Monte Carlo estimates of means and standard deviations of allocation proportions for mixed responses when $n = 60$

$E(N_{1n}/n)$	$SD(N_{1n}/n)$	$E(N_{2n}/n)$	$SD(N_{2n}/n)$	$E(N_{3n}/n)$	$SD(N_{3n}/n)$
(i) $\mu_1 = \mu_2 = \mu_3 = 2$ and $p_1 = p_2 = p_3 = 0.4$					
0.333	0.100	0.333	0.100	0.333	0.100
(ii) $\mu_1 = \mu_2 = \mu_3 = 2$, $p_1 = 0.6$, $p_2 = 0.4$ and $p_3 = 0.2$					
0.356	0.103	0.340	0.096	0.305	0.099
(iii) $\mu_1 = 3$, $\mu_2 = \mu_3 = 2$, $p_1 = 0.6$, $p_2 = 0.4$ and $p_3 = 0.4$					
0.395	0.094	0.303	0.095	0.302	0.096
(iv) $\mu_1 = 3$, $\mu_2 = \mu_3 = 2$, $p_1 = 0.6$, $p_2 = 0.4$ and $p_3 = 0.2$					
0.405	0.099	0.303	0.094	0.292	0.089
(v) $\mu_1 = 3$, $\mu_2 = 2$, $\mu_3 = 1$ and $p_1 = p_2 = p_3 = 0.4$					
0.401	0.092	0.339	0.089	0.261	0.084
(vi) $\mu_1 = 3$, $\mu_2 = 2$, $\mu_3 = 1$, $p_1 = 0.6$, $p_2 = 0.4$ and $p_3 = 0.2$					
0.415	0.089	0.335	0.095	0.250	0.082

for testing the equality of the treatment mean vectors, and suggested how the design may be adapted for mixed responses.

In the present paper, all K treatments remain in use throughout the trial, though, of course, the less promising ones are applied less often as the trial proceeds. A natural extension is to incorporate some type of elimination rule into our general model, so that the less promising treatments are dropped from further study. Work along these lines is in progress and we plan to report on this extension separately: see Biswas and Coad (2005). Similar work in the context of adaptive urn designs and binary responses has been carried out by Coad and Ivanova (2005), following earlier work by Coad (1995) for $K > 2$ treatments with normal responses.

As indicated in section 3, the choice of the w_ℓs is a subjective issue, which might have no connection with the variances. For different diseases and for different types of responses, the experimenter should set the w_ℓs at the outset depending on his/her prior belief on the relative importance of the components of the responses. Without any prior information, w_ℓ can be set to $1/m$ for all ℓ. However, the design is quite sensitive to the choice of the w_ℓs.

From a practical point of view, there are several possible extensions that would make the proposed adaptive design more flexible, especially if an elimination rule is used. For example, some of the response components may not be used in later decisions regarding which treatments to drop, and the total sample size may be adjusted in order to maintain a desirable conditional power. Clearly, such extensions

complicate the analysis considerably and are beyond the scope of the present paper.

Other extensions to the present work include the allowance for staggered entry and the consideration of generalized linear models. The latter include the logistic regression model studied by Rosenberger et al. (2001), and the adaptive survival analysis models suggested by Rosenberger and Seshaiyer (1997) and Yao and Wei (1996). Both of these extensions would make the general approach developed in this paper more widely applicable. We hope to report on extensions along these lines separately.

ACKNOWLEDGMENTS

This work was carried out while the first author was visiting the School of Mathematical Sciences, University of Sussex, during April and May 2002, and in receipt of Visiting Fellowship Research Grant GR/R92455 from the U.K. Engineering and Physical Sciences Research Council. The first author wishes to thank the School of Mathematical Sciences for its hospitality during his visit. The authors also wish to thank the three referees for their detailed comments, which have led to a much improved paper.

REFERENCES

Andersen, J., Faries, D., and Tamura, R. (1994). A Randomised Play-the-Winner Design for Multi-arm Clinical Trials, *Communications in Statistics—Theory and Methods* 23: 309–323.

Atkinson, A. C. (1982). Optimal Biased Coin Designs for Sequential Clinical Trials with Prognostic Factors, *Biometrika* 69: 61–67.

Atkinson, A. C. (1999). Optimum Biased-Coin Designs for Sequential Treatment Allocation with Covariate Information, *Statistics in Medicine* 18: 1741–1752.

Bai, Z. D., Hu, F., and Rosenberger, W. F. (2002). Asymptotic Properties of Adaptive Designs for Clinical Trials with Delayed Response, *Annals of Statistics* 30: 122–139.

Bandyopadhyay, U. and Biswas, A. (1996). Delayed Response in Randomised Play-the-Winner Rule: A Decision Theoretic Outlook, *Calcutta Statistical Association Bulletin* 46: 69–88.

Bandyopadhyay, U. and Biswas, A. (1999). Allocation by Randomised Play-the-Winner Rule in the Presence of Prognostic Factors, *Sankhyā Series B* 61: 397–412.

Bandyopadhyay, U. and Biswas, A. (2001). Adaptive Designs for Normal Responses with Prognostic Factors, *Biometrika* 88: 409–419.

Begg, C. B. and Iglewicz, B. (1980). A Treatment Allocation Procedure for Sequential Clinical Trials, *Biometrics* 36: 81–90.

Biswas, A. and Coad, D. S. (2005). Early Elimination Rule in Multi-Treatment Adaptive Allocation for Continuous Responses, unpublished.

Biswas, A. and Dewanji, A. (2004). A Randomised Longitudinal Play-the-Winner Design for Repeated Binary Data, *Australian & New Zealand Journal of Statistics* 46: 675–684.

Coad, D. S. (1995). Sequential Allocation Involving Several Treatments, in *Adaptive Designs*, N. Flournoy and W. F. Rosenberger, eds., pp. 95–109, Hayward, California: Institute of Mathematical Statistics.

Coad, D. S. and Ivanova, A. (2005). Sequential Urn Designs with Elimination for Comparing $K \geq 3$ Treatments, *Statistics in Medicine*, in press.

Holmes, L. B., Harvey, E. A., Brown, K. S., and Khosbin, S. (1994). Anticonvulsant Teratogenesis: 1, A Study Design for Newborn Infants, *Teratology* 49: 202–207.

Ivanova, A. and Rosenberger, W. F. (2000). A Comparison of Urn Designs for Randomised Clinical Trials of $K > 2$ Treatments, *Journal of Biopharmaceutical Statistics* 10: 93–107.

Ivanova, A., Rosenberger, W. F., Durham, S. D., and Flournoy, N. (2000). A Birth and Death Urn for Randomised Clinical Trials: Asymptotic Results, *Sankhyā Series B* 62: 104–118.

Jennison, C. and Turnbull, B. W. (1993). Group Sequential Tests for Bivariate Response: Interim Analyses of Clinical Trials with both Efficacy and Safety Endpoints, *Biometrics* 49: 741–752.

Jennison, C. and Turnbull, B. W. (2000). *Group Sequential Methods with Applications to Clinical Trials*, London: Chapman and Hall.

Little, R. J. and Schluchter, M. D. (1985). Maximum Likelihood Estimation for Mixed Continuous and Categorical Data with Missing Values, *Biometrika* 72: 497–512.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979). *Multivariate Analysis*, London: Academic Press.

Rosenberger, W. F. (1993). Asymptotic Inference with Response-Adaptive Treatment Allocation Designs, *Annals of Statistics* 21: 2098–2107.

Rosenberger, W. F. and Seshaiyer, P. (1997). Adaptive Survival Trials, *Journal of Biopharmaceutical Statistics* 7: 617–624.

Rosenberger, W. F., Vidyashankar, A. N., and Agarwal, D. K. (2001). Covariate-Adjusted Response-Adaptive Designs for Binary Response, *Journal of Biopharmaceutical Statistics* 11: 227–236.

Royall, R. M. (1991). Ethics and Statistics in Randomised Clinical Trials (with Discussion), *Statistical Science* 6: 52–88.

Tamura, R. M., Faries, D. E., Andersen, J. S., and Heiligenstein, J. H. (1994). A Case Study of an Adaptive Clinical Trial in the Treatment of Out-Patients with Depressive Disorder, *Journal of the American Statistical Association* 89: 768–776.

Wang, X., Wypij, D., Gold, D. R., Speizer, F. E., Ware, J. H., Ferris, B. G., and Dockery, D. W. (1994). A Longitudinal Study of the Effects of Parental Smoking on Pulmonary Function in Children 6–18 Years, *American Journal of Respiratory and Critical Care Medicine* 149: 1420–1425.

Wei, L. J. and Durham, S. D. (1978). The Randomised Play-the-Winner Rule in Medical Trials, *Journal of the American Statistical Association* 73: 840–843.

Whitehead, J. (1997). *The Design and Analysis of Sequential Clinical Trials*, rev. Second edition, Chichester, England: Wiley.

Yao, Q. and Wei, L. J. (1996). Play the Winner for Phase II/III Clinical Trials, *Statistics in Medicine* 15: 2413-2423.